ARMY RESEARCH LABORATORY

Implementation of Efficient Pan-Tilt-Zoom Camera Calibration

by Nicholas Fung and Philip David

ARL-TR-4799 **April 2009**

Approved for public release; distribution unlimited.

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Adelphi, MD 20783-1197

ARL-TR-4799 **April 2009**

Implementation of Efficient Pan-Tilt-Zoom Camera Calibration

Nicholas Fung and Philip David
Computational and Information Sciences Directorate, ARL

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
April 2009	Final	October 2007 to August 2008

4. TITLE AND SUBTITLE
Implementation of Efficient Pan-Tilt-Zoom Camera Calibration

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Nicholas Fung and Philip David

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: AMSRD-ARL-CI-IA
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-4799

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Pan-tilt-zoom (PTZ) cameras, frequently used in both online and automated surveillance applications, require accurate knowledge of camera parameters in order to accurately register autonomously tracked objects to a world model. Due to imprecision in the PTZ mechanism, these parameters cannot be obtained from PTZ control commands but must be calculated directly from camera imagery. This report describes the efforts to implement a real-time calibration system for a stationary PTZ camera. The approach continuously tracks distinctive image feature points from frame to frame, and from these correspondences, robustly calculates the homography transformation between frames. Camera internal parameters are then calculated from these homographies. Finally, the external parameters can be calculated from the internal parameters and image homographies. The calculations are performed by a self-contained program that continually monitors images collected by the camera as it performs pan, tilt, and zoom operations. The accuracy of the calculated calibration parameters are compared to ground truth data. The program works with a higher degree of accuracy for small changes in the camera's external parameters. In addition, long algorithm execution time prevents the algorithm from running under all real-time conditions.

15. SUBJECT TERMS
Camera calibration, PTZ camera, image homography, camera parameters

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Nicholas Fung
Unclassified	Unclassified	Unclassified	UU	22	19b. TELEPHONE NUMBER (*Include area code*) (301) 394-3101

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Introduction

As technology has advanced, the potential for security and surveillance systems has increased. Better communications technology allows for large amounts of data to stream quickly throughout a network. Faster processors allow the use of advanced algorithms to process the data. Potentially, large numbers of both stationary and mobile cameras can be used cooperatively to form a persistent surveillance system to provide visual coverage of a large area. Algorithms to detect, track, and place targets in a real-world model would make an impressive system. However, the practical use of such a system is very dependent upon the ability to acquire calibrated images from the camera with a high degree of accuracy. Inaccurate knowledge of camera parameters can lead to missed or misidentified targets, in addition to errors in locating the targets within a world model. As such, camera calibration is an important subject.

Proper camera calibration is important in establishing accurate images and camera positioning to perform other visual surveillance tasks, such as target tracking and image mosaicking. The calibration matrix consists of five values: the aspect ratio, the skew value, the focal length, and the x and y values of the principal point. These parameters are needed to establish the position of targets within a world model. They are also needed when manipulating several camera images, such as when creating a mosaic of the observed area. In addition to determining the camera's internal parameters, calculations can determine a rotation matrix, allowing for pan and tilt values to be established for the camera. The camera image can also suffer from both radial and tangential distortion, particularly at extreme zoom values. If the distortions can be properly corrected, it can lead to more accurate target location and image mapping.

As part of the U.S. Army Research Laboratory's (ARL) goal of integrating various mobile and stationary sensor assets into a comprehensive persistent surveillance system, proper camera calibration is vital. This report documents steps taken toward achieving a real-time calibration routine with a high degree of accuracy. While camera calibration parameters can be computed offline using a series of images, real-time calibration will allow a higher degree of accuracy as the routine can continually update the calibration parameters. This capability is important, because the calibration can change over time and movement, particularly when performing zoom functions.

2. Self Calibration

The calibration matrix is a 3x3 matrix that incorporates the camera's intrinsic calibration parameters. The calibration matrix of frame *k* is

$$K_k = \begin{bmatrix} \gamma_k f_k & s_k & x_k \\ 0 & f_k & y_k \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

where γ_k is the aspect ratio, f_k is the focal length, s_k is the axis skew, and x_k and y_k are the x and y values of the principal point, respectively. The aspect ratio is a value that relates the horizontal pixel length to the vertical pixel length. For a camera with equal pixel width and height, the aspect ratio is one, as is the case in many modern cameras. The focal length is directly related to the zoom of the camera. A larger focal length indicates a higher zoom. The skew value indicates a misalignment between the camera axes, resulting in a slanted image. The principal point is the point in the image through which the optical axis passes. Other factors that influence the captured images involve radial and tangential distortions. These distortions act upon the image to alter how it is captured and are not represented in the calibration matrix. Radial distortion consists of barrel distortions, which are large with wide angle or low zoom lenses, and pincushion distortions, which are large for telephoto or large zoom lenses. Tangential distortions occur under imperfect centering of lens components. Such distortions can be corrected for using an appropriate camera model (*4*).

There are several methods for calibrating a camera. For a camera capable of pan, tilt, and zoom (PTZ) operations, multiple images can be captured at different orientations. The images can be analyzed to calculate the inter-image homography between each pair using corresponding point pairs. The homography is a matrix that relates corresponding points from one image to another. In the context of a PTZ camera, the homography is a matrix that allows one image to be aligned to the next and form a mosaic. The homography can be analyzed to establish camera calibration parameters between each pair of images.

Several algorithms can be used to determine the point correspondences between two images. Two of the most popular methods are the Kanade-Lucas-Tomasi (KLT) Feature detector and the Scale-Invariant Feature Transform (SIFT). For this project, we chose SIFT for the initial approach to the problem, because the algorithm is invariant to translation, scaling, and rotation, in addition to being partially invariant to illumination changes and affine projections (*1*). The robustness of the algorithm would be appropriate for handling changes captured in the cameras. For example, the algorithm would be robust against an illumination change resulting from a cloud passing in front of the sun. After matching points between the current and previous image, the homography can be computed. After some analysis, we employed a direct linear transform

(DLT) algorithm with normalization to perform this calculation. Other options were a nonlinear algorithm and a DLT without normalization. We evaluated these approaches by creating ideal corresponding points using known pan, tilt, zoom, and principal point settings. The correspondences were then perturbed by a set amount and the resulting pan, tilt, zoom, and principal point values were calculated. The results can be seen graphically in figure 1. The graphs compare the results between the unnormalized and normalized linear approaches. For each of these algorithms, three lines are plotted: the minimum error, the maximum error, and the median error. Each of the parameters showed a sharp increase as the error in point correspondences increased from 0 to 0.1 pixels. The errors then increased at a slower rate with higher point correspondence errors. The normalized linear method performed with the lowest overall error. This result is particularly evident with the maximum error. The nonlinear algorithm is not depicted in the figure as it was determined to be too slow to be useable in a real-time system.

3. Approach

3.1 Rooftop Camera System

The calibration routine was created for and tested on the rooftop camera system currently in place at the ARL Adelphi, MD, building. The system includes four Sony Network Cameras (SNCs) (Sony #SNC-RZ30) positioned at each corner of the building. One of the features of the SNC is the ability to feed PTZ values to modify the orientation and zoom of the camera. In addition, these values can be read back. While these features would seem to make calibrating the extrinsic parameters unnecessary, the program should still perform this task so that other cameras can be calibrated.

The cameras are all connected to an internal network and imagery can be streamed to or captured by any computer on the network. We are in the process of creating programs that process the data and perform functions such as target tracking and image mosaicking. Among such programs, a procedure to perform camera calibration is important to establish accurate data.

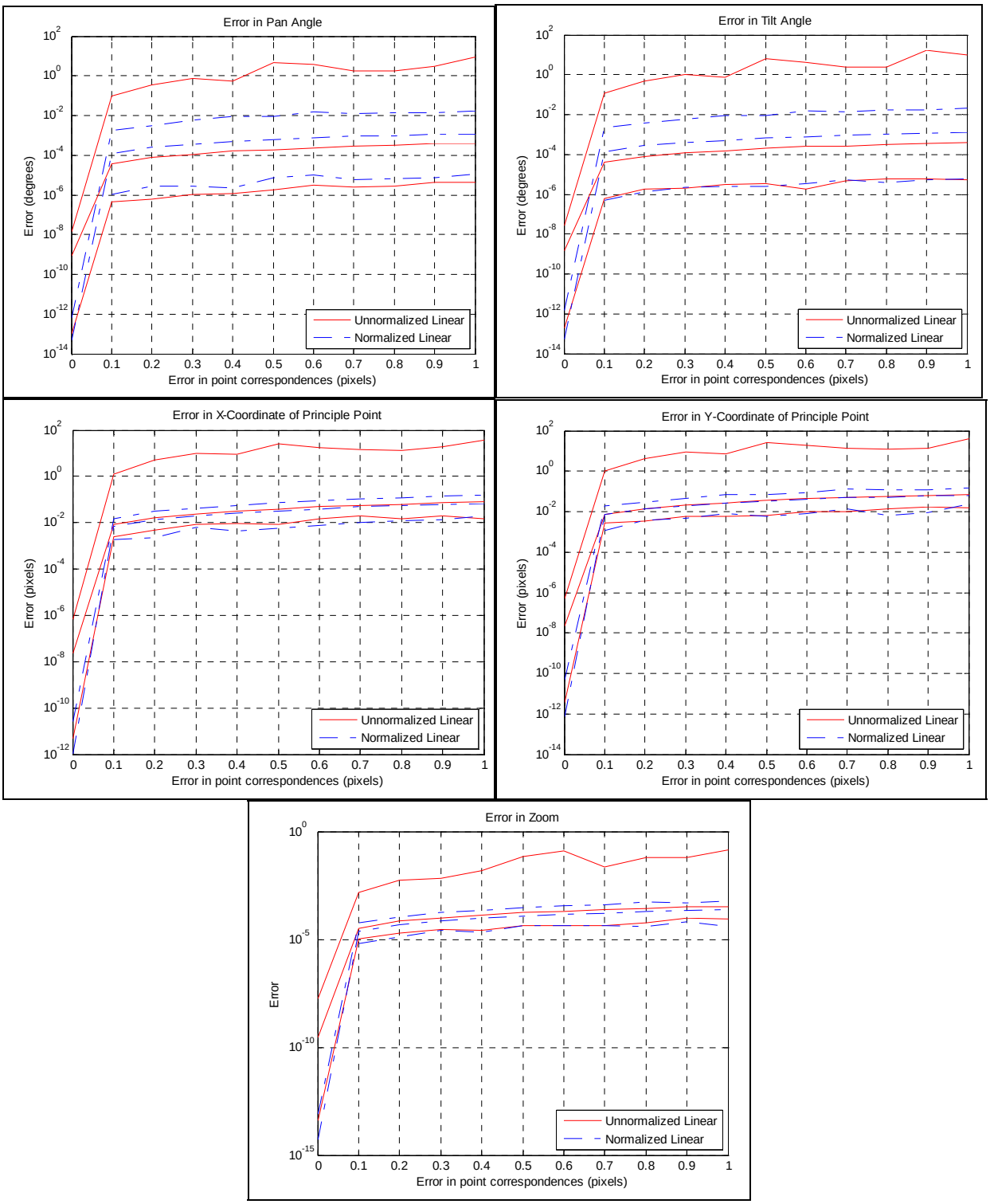

Figure 1. Testing of homography calculation algorithms under corresponding point errors.

3.2 Calculations

To establish baseline calibration parameters, we used a MATLAB code created by Bouguet of the California Institute of Technology (*5*). The toolbox implements a derivative of the method developed by Heikkil and Silven at the University of Oulu in Finland (*3*). This method uses the DLT to solve for the projection of an object onto the image, thus implicitly obtaining the calibration parameters. The algorithm requires that a test pattern, resembling a checkerboard, be captured by the camera in a series of images. The operator aligns the pattern to the image by manually locating the corners of the grid within the image. Because the test pattern is of known measurements, the projection can be calculated. Solving for the projection yields the calibration parameters.

Once an initial calibration matrix has been solved, further calibrating can be performed more easily on successive frames without the time-consuming process of capturing the test pattern. First, we calculate the initial image of the absolute conic (IAC), ω_0, based on the initial calibration matrix:

$$\omega_0 = \left(K_0 K_0^{\ T}\right)^{-1}. \tag{2}$$

We calculate the next IAC using the following formula. This equation is an application of the Kruppa equation 6:

$$\omega_i = \left(H_i\right)^{-T} \omega_{i-1} \left(H_i\right)^{-1}. \tag{3}$$

We derive the new calibration matrix, K_i, using the Cholesky Decomposition of the IAC. The Cholesky Decomposition returns two matrices, one lower and one upper triangular matrix, whose product is the input:

$$K_i = chol(\omega_i). \tag{4}$$

Finally, we compute the rotation matrix, R_i, using the following formula:

$$R_i = K_i^{\ -1} H_i K_{i-1}. \tag{5}$$

From the rotation matrix, we can calculate pan and tilt values:

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}$$

$$tilt = \tan^{-1}\left(\frac{-R_{23}}{R_{33}}\right)$$

$$pan = \sin^{-1}\left(R_{13}\right)$$

$$roll = \tan^{-1}\left(\frac{-R_{12}}{R_{11}}\right). \tag{6}$$

These values record the change of the camera orientation in that direction. We can calculate the absolute zoom using the focal length values of the calibration matrix:

$$zoom = -\frac{K_i(1,1) + K_i(2,2)}{2K_1(1,1)}. \tag{7}$$

With the new IAC, we can readily repeat the process. Each frame just requires the few simple calculations, which are relatively quick to calculate. In addition, changes in the pan, tilt, and zoom values can be accumulated to record the orientation of the camera.

To increase the accuracy of the procedure, we have to account for and correct the distortion values. Barrel distortions are more prevalent at low zoom settings, and pincushion distortions are more prevalent at high zoom settings. In the case of the Sony cameras used for testing, the barrel distortions seemed noticeable at the lowest zooms, while the pincushion distortions were not as noticeable at high zoom settings. In addition to the parameters of the calibration matrix, the MATLAB code created by Bouguet also calculates five distortion parameters, c_i, $i = 1,\ldots,5$. The lens-distorted pixel location of a normalized (undistorted, true pinhole camera) image point $x = (x,y)^T$ is $x_d = (x_d,y_d)^T$, where

$$\mathbf{x}_d = \begin{bmatrix} x_d \\ y_d \end{bmatrix} = (1 + c_1 r^2 + c_2 r^4 + c_5 r^6)\mathbf{x} + \mathbf{d}_x, \tag{8}$$

$$\mathbf{d}_x = \begin{bmatrix} 2c_3 xy + c_4(r^2 + 2x^2) \\ c_4(r^2 + 2y^2) + 2c_4 xy \end{bmatrix}. \tag{9}$$

The term $\mathbf{d}_x$ represents the tangential distortion correction. The variable r represents the distance of the point $\mathbf{x}$ from the center of the image. The homography can be calculated from the feature points found in these undistorted images. By correcting for these distortions, the homography will be more accurate and thus subsequent calculations to find the calibration parameters will be more accurate.

4. Experiments

We encountered several difficulties during the creation of the camera calibration routine. One of the first problems was in finding and implementing a viable algorithm both in processing time and accuracy. The first approach toward achieving calibration involved applying a number of assumptions to simplify the problem. We assumed the skew value to be zero, the aspect ratio to be one, and the principal point to be centered. This approach was detailed in reference 2. We made these assumptions, because modern, high-quality cameras tend to have attributes close to these values. However, after some testing with this algorithm, we found the accuracy broke down quickly, particularly under zoom changes and moderate angle changes. We determined these inaccuracies were unacceptable, so we reworked the process to avoid making a large number of assumptions. Further testing showed that the calibration parameters changed and could not be assumed to be a specific value. In particular, the principal point changed greatly under extreme zoom functions. The principal point shift can be seen graphically in figure 2.

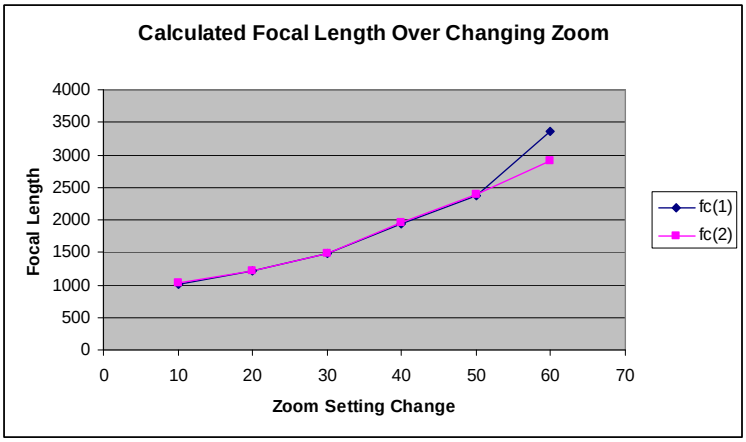

Figure 2. Focal length as calculated through the calibration routine.

Another problem we encountered was with the current implementation of the camera calibration algorithm. The program obtained good results under moderate to small changes in camera orientation and zoom values. However, the accuracy broke down as the changes got larger. This problem could have arisen due to inaccuracies in the homography that was used to calculate the calibration parameters. To work against this problem, the calibration calculations can be taken quickly and often to calibrate under small PTZ conditions.

The largest problem we encountered was in calculating the homography. An accurate homography is vital to obtaining accurate calibration parameters. For this reason, we employed the SIFT method. SIFT's ability to perform under scale changes is important for calibrating cameras that will be performing zoom operations. However, the implementation of SIFT used

for the calibration routine proved to be time consuming. To combat this problem, we calculated fewer points, sacrificing accuracy for expediency. However, this method still did not produce the speed necessary for a real-time system. Other alternatives not yet implemented and tested could include using a faster processor or a dedicated graphical processing unit (GPU) (*7*).

Because the distortion values are difficult to calculate on the fly, we calculated the values beforehand using the Bouguet MATLAB code. The values were stored in a database to be used as the camera underwent zoom functions. The captured images were corrected for distortion before being processed to calculate the homography and calibration values. Figure 3 shows the changes in the distortion parameters graphically. The camera returns a value between 0 and 100 to indicate the amount of camera zoom. Figure 4 provides an interpretation of this parameter. The parameters stay relatively small for zooms under 50. However, they begin to vary widely as the zoom is further increased. The fifth distortion parameter was zero under all measured zooms and is not shown.

We computed the camera focal lengths at each step-of-10 increase in zoom using the Bouguet MATLAB code. Figure 4 shows those results. We also calculated the aspect ratio. The results show that the aspect ratio stays close to one throughout the various zoom settings. The focal length changes at a rate that resembles exponential growth.

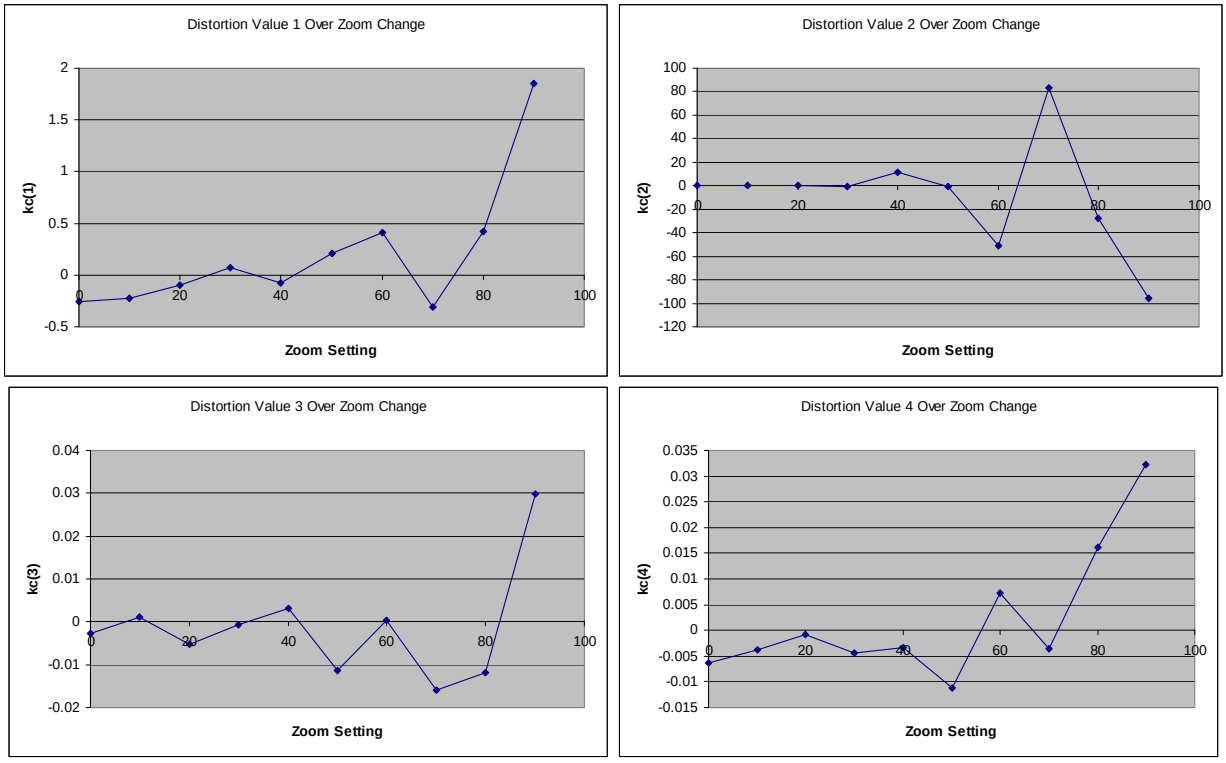

Figure 3. Distortion parameter changes.

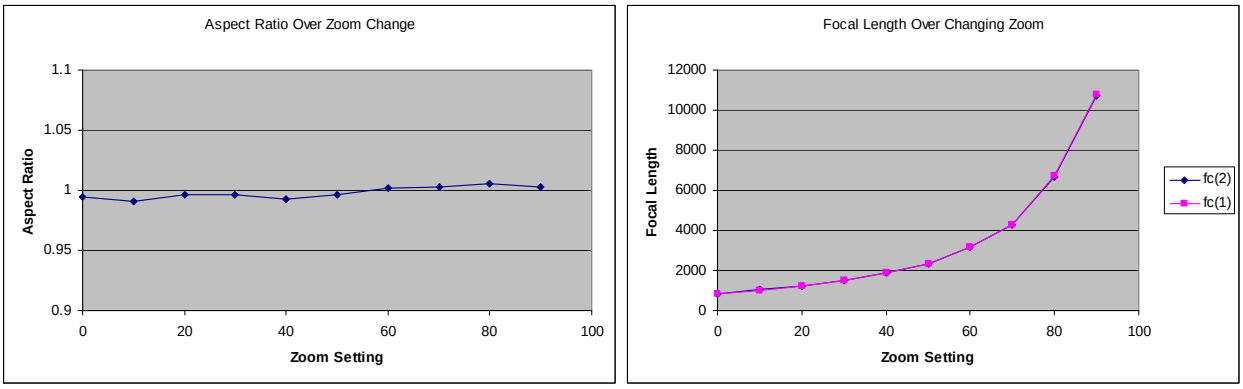

Figure 4. Aspect ratio and focal length changes over a changing zoom.

We then used the calibration routine to calculate the focal length according to a series of images with a changing zoom. We calibrated each zoom image from the initial image of zoom 0. For example, the image of zoom 40 was calibrated using the homography relating it to the image of zoom 0 and the result was approximately 2000. The total results appear graphically in figure 3. The calculated values follow the results from figure 4 closely through small zoom changes. However, at a zoom change of 60, the vertical focal length began to diverge from the horizontal focal length. We did not calculate zoom changes of over 70, because there were not enough corresponding points to calculate a homography. Again, the zoom values used are built in zoom settings of the camera from a 0 to 100 scale.

We also recorded the shift of the principal point over changes in zoom. The results are shown in figure 5. As with the distortion parameters, the principal point is steady under low zooms, but starts to shift greatly with higher zoom settings. Under ideal conditions, the principal point would be the exact center of the image. In the case of this camera capturing a 640x480 resolution image, we would estimate the principal point to be 320 horizontal and 240 vertical. At low zoom settings, the calibration routine returns numbers similar to these expected values. At higher zoom settings, the principal point shifts to a higher degree. It is noteworthy that the horizontal and vertical coordinates follow a similar shift pattern as the zoom setting increases.

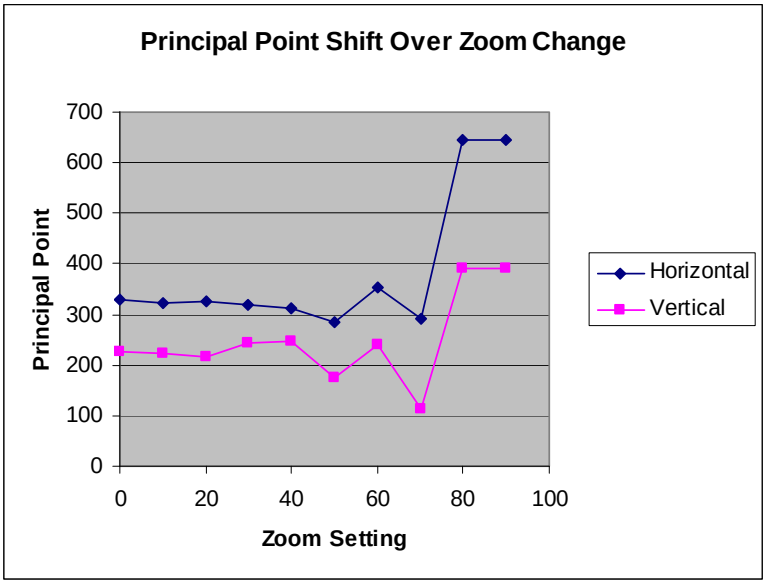

Figure 5. Principal point shift over a changing zoom.

The final test was applying the algorithm to an application. Figure 6 shows an image mosaic created with the calibrated data. We created the mosaic using 25 images at different pan and tilt values. We held the zoom steady at the widest field of view. The figure shows good alignment between the images as they are placed into the mosaic, indicating good calibration at this zoom setting.

Figure 6. Example of a mosaic created with calibration values.

We created the mosaic in figure 6 using images taken from pan values of –25 to –45 and tilt values of 10 to 30, each in 5° increments. Figure 7 graphically depicts the change in pan and tilt values returned by the calibration algorithm against the values returned by the camera itself. The top graph shows pan values while holding the tilt at five different constant values. The bottom graph shows tilt values at each of five different pan values. The points would be expected to follow a linear, $x = y$ equation. However, the growth is steeper, particularly at high pan and tilt values.

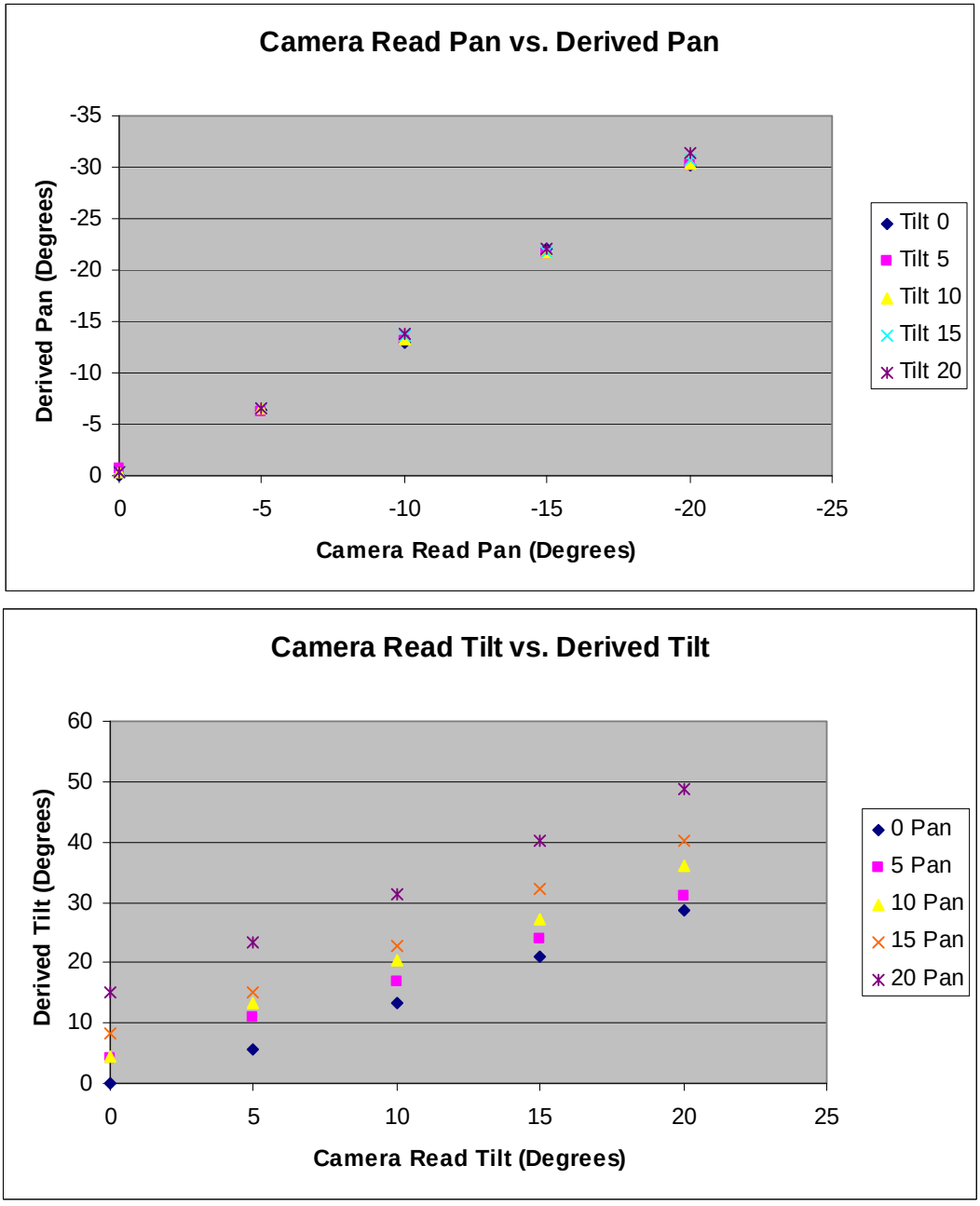

Figure 7. Derived pan/tilt values against camera read values.

This result suggests that the algorithm breaks down as the extrinsic value changes increase in magnitude. In addition, tilting the camera did not influence the pan values as much as panning influenced the tilt values. The first graph shows that the pan values matched well, even at different tilt values. The second graph shows that a higher pan created a higher tilt value. There are several possible explanations for this phenomenon. One possibility is that the algorithm breaks down with larger pan angles. Another possibility is an inconsistent camera pan/tilt mechanism. Further study is needed to test this phenomenon.

6. Conclusions

The camera calibration routine is accurate for small PTZ operations. However, as the change in the camera orientation grows larger, the accuracy breaks down. This conclusion is particularly evident for calculated tilt angles under a large change in the pan angle. The accuracy breakdown could be the result of inaccurate homographies, inaccuracies of the equations, or imprecise movement of the camera PTZ mechanism. Further work needs to be done to improve the performance, including the possibility of using a different feature tracker. Alternatively, calibration could be performed at a frequent rate to avoid having to calibrate over a large pan, tilt, or zoom operation. This method should improve the performance, because the data has shown better accuracy over small PTZ operations. Unless the camera is changing at a fast rate, frequent calibrations should have a small chance of calibrating over a large orientation change.

A limiting characteristic of the procedure is the speed of the SIFT algorithm. SIFT proved to be accurate and has the positive characteristic of being robust to scale changes; however, the implementation used required a large amount of processing time. As a result, the program cannot run under normal real-time scenarios. Improvements can be made in terms of processor speed or by assigning a designated processor, such as a GPU, to perform the task. Alternatively, we are exploring other implementations of SIFT and other methods of obtaining a homography.

7. References

1. Lowe, D. Object Recognition from Local Scale-Invariant Features. *Proceedings of the 7th IEEE International Conference on Computer Vision*, Kerkyra, Greece, Sept 1999, 1150–1157, vol. 2.

2. Kim, J.; Hong, K. S. A Practical Self-Calibration Method of Rotating and Zooming Cameras. *IEEE* **2000**, 354–357.

3. Heikkila, J.; Silven, O. A Four-step Camera Calibration Procedure with Implicit Image Correction. *Proceedings from the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 17–19 June 1997, 1106–1112.

4. Brown, D. C. Close-Range Camera Calibration. *Photogrammetric Engineering* **1971**, *37* (8), 855–866.

5. Bouguet, J.-Y. *Camera Calibration Toolbox for Matlab*. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html, 2 June 2008.

6. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004.

7. Heymann, S.; Müller, K.; Smolic, A.; Fröhlich, B.; Wiegand, T. SIFT Implementation and Optimization for General-purpose GPU. *15th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen, Czech Republic, 29 Jan–1 Feb 2007, 317–322.

List of Symbols, Abbreviations, and Acronyms

ARL	U.S. Army Research Laboratory
DLT	direct linear transform
GPU	graphical processing unit
IAC	image of the absolute conic
KLT	Kanade-Lucas-Tomasi Feature
PTZ	pan, tilt, and zoom
SIFT	Scale-Invariant Feature Transform
SNCs	Sony Network Cameras

No. of Copies	Organization
1 ELEC	ADMNSTR DEFNS TECHL INFO CTR ATTN DTIC OCP 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	DARPA ATTN IXO S WELBY 3701 N FAIRFAX DR ARLINGTON VA 22203-1714
1 CD	OFC OF THE SECY OF DEFNS ATTN ODDRE (R&AT) THE PENTAGON WASHINGTON DC 20301-3080
1	US ARMY RSRCH DEV AND ENGRG CMND ARMAMENT RSRCH DEV AND ENGRG CTR ARMAMENT ENGRG AND TECHNLGY CTR ATTN AMSRD AAR AEF T J MATTS BLDG 305 ABERDEEN PROVING GROUND MD 21005-5001
1	PM TIMS, PROFILER (MMS-P) AN/TMQ-52 ATTN B GRIFFIES BUILDING 563 FT MONMOUTH NJ 07703
1	US ARMY INFO SYS ENGRG CMND ATTN AMSEL IE TD F JENIA FT HUACHUCA AZ 85613-5300
1	COMMANDER US ARMY RDECOM ATTN AMSRD AMR W C MCCORKLE 5400 FOWLER RD REDSTONE ARSENAL AL 35898-5000
1	US GOVERNMENT PRINT OFF DEPOSITORY RECEIVING SECTION ATTN MAIL STOP IDAD J TATE 732 NORTH CAPITOL ST NW WASHINGTON DC 20402

No. of Copies	Organization
1	US ARMY RSRCH LAB ATTN AMSRD ARL CI OK TP TECHL LIB T LANDFRIED BLDG 4600 ABERDEEN PROVING GROUND MD 21005-5066
1	DIRECTOR US ARMY RSRCH LAB ATTN AMSRD ARL RO EV W D BACH PO BOX 12211 RESEARCH TRIANGLE PARK NC 27709
6	US ARMY RSRCH LAB ATTN AMSRD ARL CI IA N FUNG ATTN AMSRD ARL CI IA P DAVID ATTN AMSRD ARL CI IA S YOUNG ATTN AMSRD ARL CI OK PE TECHL PUB ATTN AMSRD ARL CI OK TL TECHL LIB ATTN IMNE ALC HR MAIL & RECORDS MGMT ADELPHI MD 20783-1197
TOTAL:	16 (1 ELEC, 1 CD, 14 HCs)

INTENTIONALLY LEFT BLANK.